PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/12, C07K 14/47, C12Q 1/68, A61K 38/17, C07K 16/18

A2

(11) International Publication Number: **WO 00/08156**

(43) International Publication Date: 17 February 2000 (17.02.00)

(21) International Application Number: PCT/US99/17800

(22) International Filing Date: 6 August 1999 (06.08.99)

(30) Priority Data:

60/155,245	7 August 1998 (07.08.98)	US
60/155,181	27 August 1998 (27.08.98)	US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications

US	60/155,181 (CIP)
Filed on	27 August 1998 (27.08.98)
US	60/155,245 (CIP)
Filed on	7 August 1998 (07.08.98)

(71) Applicant *(for all designated States except US)*: INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). REDDY, Roopa [IN/US]; 1233 W. McKinley Drive, Sunnyvale, CA 94086 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US).

(74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: DNA REPLICATION- AND REPAIR-ASSOCIATED PROTEINS

(57) Abstract

The invention provides human DNA replication- and repair-associated proteins (DRASP) and polynucleotides which identify and encode DRASP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating or preventing disorders associated with expression of DRASP.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

DNA REPLICATION- AND REPAIR-ASSOCIATED PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of DNA replication- and repair-associated proteins and to the use of these sequences in the diagnosis, treatment, and prevention of developmental disorders, proliferative disorders including cancer, and autoimmune/inflammatory disorders.

BACKGROUND OF THE INVENTION

Replication of individual cells is fundamental to the survival of multicellular organisms. Cells replicate as an organism grows and are replaced in response to normal aging as well as to destruction by trauma or infection. Cells conserve genetic material by replicating DNA. Somatic cells replicate by mitosis. In this process, the DNA in the cell nucleus is uncoiled, and the two halves of each duplex molecule are replicated. The identical paired DNA strands migrate to separate locations and are incorporated into the two nuclei of two daughter cells. Meiosis produces egg and sperm cells, which divide and migrate to produce daughter cells that have half the parent cell chromosome count. A variety of cell structures, regulatory mechanisms and reaction pathways accomplish cell replication with great speed and precision despite the physical and chemical complexity of the uncoiling, replication, division, migration and synthesis of new cellular materials. Yet other mechanisms evolved to correct errors or damage incurred during replication and division. Cells may also undergo metastasis (cancerous growth) or apoptosis (controlled destruction).

Annexin I (also called lipocortin I) is a calcium/phospholipid-binding protein which promotes membrane fusion and is involved in exocytosis. Annexin I has been shown to play a role in cell proliferation, differentiation, and apoptosis. (Kim et al. (1997) Biochem Mol. Biol. Int. 43:521-528; Solito et al. (1998) Cell Growth Differ. 9:327-336.) This protein appears to bind between two and four calcium ions with high affinity and contains four homologous repeat amino acid sequences which are common to all annexin proteins. A pair of these repeats have been suggested to form one binding site for calcium and phospholipid. (SWISS-PROT:P04083.) Annexin I regulates the activity of phospholipase A2, an enzyme which plays a critical role in inducing mitosis. (Kim et al. (1998) Mol. Cells 8:90-95.) Annexin I levels increase during recovery from renal ischemia, a process that involves both hyperplasia and hypertrophy. (McKanna et al. (1992) J. Cell. Physiol. 153:467-476.) Annexin I levels are increased in proliferative hepatocytes, and annexin has been suggested as a useful marker for distinguishing actively proliferating, pre-proliferating, and post-proliferative hepatocytes. (Masaki et

al. (1994) Hepatology 20:425-435.) In addition, annexin I levels may correlate with malignant breast cancer progression. (Ahn et al. (1997) Clin. Exp. Metastasis 15:151-156.) Treatment with okadaic acid, which increases the level of annexin I in a promyelocytic leukemia cell line, causes the leukemia cells to differentiate into macrophage-like cells rather than to undergo mitosis. (Sato et al. (1995) Biochim. Biophys. Acta 1266:23-30.)

DNA may be damaged by proliferation. Double-strand breaks and single-strand gaps in damaged DNA are repaired by mechanisms associated with a series of complex biochemical reactions known as recombination. A variety of genes associated with recombination have been identified in prokaryotes, and these genes are highly conserved. Genes homologous to bacterial and yeast recombination genes have been identified in higher eukaryotes. (Shinohara et al. (1995) Trends Biochem. Sci. 20:387-391.)

The prokaryotic protein recF is a single-stranded DNA-binding protein which also probably binds ATP. RecF is required for recombinational DNA repair. (Howard-Flanders, F. (1975) Basic Life Sci. 5A:265-274.) E. coli, for example, may carry out homologous genetic recombination by multiple pathways, and RecF may act as a switch between replication and recombination by helping to convert replication to recombination intermediates. (Clark et al. (1994) Crit. Rev. Microbiol. 20:125-142.) Genes shown to encode RecF proteins show relatively little sequence homology across species. (Roberts et al. (1997) J. Bacteriol. 179:2319-2330; Gorbalenya et al. (1990) J. Mol. Biol. 213:583-591.) Even a small degree of homology to known RecF motifs may therefore be regarded as significant.

Cells are constantly faced with replication errors and environmental assault (such as ultraviolet irradiation) that can produce DNA damage. Damage to DNA consists of any change that modifies the structure of the molecule. Changes to DNA can be divided into two general classes, single base changes and structural distortions. Single base changes affect the sequence but not the overall structure of the DNA. Since single base changes do not affect transcription or replication, they exert their effect on future generations. Structural distortions affect the structure of the DNA. A single strand nick or removal of a base may prevent a strand from acting as a viable template for synthesis of DNA or RNA. Intrastrand or interstrand covalent linkage between bases, or the addition of a bulky adduct to a base, may distort the structure of the double helix and interfere with transcription and replication. Any damage to DNA can produce a mutation, and the mutation may produce a disorder, such as cancer.

Changes in DNA are recognized by repair systems within the cell. These repair systems act to correct the damage and thus prevent any deleterious affects of a mutational event. Repair systems can be divided into three general types, direct repair, excision repair, and retrieval systems. When the repair systems are eliminated, cells become exceedingly sensitive to environmental mutagens, such as ultraviolet irradiation. Disorders associated with a loss in DNA repair systems often exhibit a high

sensitivity to environmental mutagens. Examples of such disorders include xeroderma pigmentosum (XP), Bloom's syndrome (BS), and Werner's syndrome (WS). (Yamagata, K., et al. (1998) Proc. Natl. Acad. Sci. USA 95:8733-8738.)

Direct repair involves the reversal or simple removal of the damaged region of DNA. Mismatches involving normal bases are repaired based on certain biases within the repair system. For example, mismatched G•T base pairs are frequently caused by deamination of 5-methyl-cytosine to form thymine. Therefore, repair systems convert mismatched G•T pairs to G•C, instead of A•T. Repair also favors the non-methylated strand in hemimethylated DNA, since this strand represents the newly synthesized daughter strand. The recognition of hemimethylated DNA and repair of mismatches on the non-methylated strand involve the products of the genes *mutH*, *mutL*, *mutS* (which specifically recognizes mismatched base pairs), the helicase encoded by the *uvrD* gene, and the methylase encoded by the *dam* gene.

Excision repair is a system in which mispaired or damaged bases are removed from DNA and a new stretch of DNA is synthesized to replace them. In the incision step, the damaged structure is recognized by an endonuclease that cleaves the DNA strand on both sides of the damage. In the excision step, a 5'-3' exonuclease removes a stretch of the damaged DNA strand. In the synthesis step, the resulting single-stranded region serves as a template for a DNA polymerase to synthesize a replacement for the excised sequence. Finally, DNA ligase covalently links the 3' end of the new material to the old material.

Although eukaryotic DNA repair systems have been described, they are primarily based on homology with the well described prokaryotic DNA repair processes. In prokaryotes, two types of excision repair exist. Short-patch repair (SPR) removes 20 nucleotides on average. SPR accounts for 99% of excision-repair events and acts as a constitutive function of the cell. Long-patch repair (LPR) removes 1500 nucleotides on average, but can remove more than 9,000 bases. LPR accounts for the remaining 1% of excision repair events and is induced by DNA damage. Both short- and long-patch repair involve the *uvr* genes *uvrA*, *uvrB*, and *uvrC* that code for the components of a repair endonuclease. The endonuclease recognizes bulky lesions, such as pyrimidine dimers, and makes an incision 7 nucleotides from the 5' side of the damaged site and ~3 nucleotides from the 3' side. Several E. coli enzymes can excise bulky lesions from DNA in which such excisions have been made, including the 5'-3' exonuclease activities of DNA PolI and PolII, and single strand specific exonuclease VII. DNA PolI is thought to perform most of the excision repair *in vivo*. The helicase activity of *uvrD* is also needed.

If replication occurs prior to repair of bulky lesions in the DNA template, DNA polymerase will cease synthesis of the daughter strand close to the lesion and resume replication some distance

further along the template. This process leaves a substantial gap in the newly synthesized strand. The other parental strand (the complementary strand), which does not contain the lesion, produces a normal daughter strand. Repair of the gap involves "stealing" the homologous single strand of DNA from the normal duplex. Following this single strand exchange, the recipient duplex has a parental (damaged) strand facing a wild-type strand. The donor duplex has a normal parental strand facing a gap. This gap can be filled by normal replication machinery, generating a normal duplex. This system does not repair damage, but confines the damage to the original distortion. The same recombination-repair events must be repeated after every replication cycle unless the damage is actually removed by an excision-repair system. The genes involved in retrieval repair include *recB*, *recC*, *recF*, *recQ*, and *recA*. These genes fall into two recombination-repair pathways: *recA* is a component of both pathways, while *recBC* is a component of one and *recFQ* is a component of the other. The *recBC* genes code for two subunits of exonuclease V. In *recA* mutants exonuclease V is uncontrolled and causes extensive DNA digestion. The *recQ* gene codes for a helicase capable of unwinding a wide variety of DNA substrates, including joint molecules formed by RecA protein. (Harmon, F.G., and Kowalczydowski, S.C. (1998) Gene Dev. 12:1134-1144.) The function of the *recF* gene product is unknown.

The RecA protein has two distinct functions. The first function involves invasion of a single DNA strand into a DNA duplex. This function requires RecA to have sites responsible for DNA binding, and this activity is directly involved in the Retrieval System as well as in general recombination events. The second function involves activation of gene expression. RecA is activated by treatments that block DNA replication (like ultraviolet irradiation). Once activated, RecA triggers latent proteolytic activities of target proteins, and the resulting proteolysis targets gene regulatory proteins which suppress transcription at specific operons. The activation of RecA is responsible for inducing the expression of many genes, whose products include repair functions including, for example, the long-patch repair mode.

Eukaryotic homologs to prokaryotic DNA repair enzymes have been described. The Rad51 family of proteins are homologs of RecA. (Baumann, P., and West, S.C. (1998) Trends Biochem. Sci. 23:247-251.) Yeast Sgs1, and human BLN and WRN are homologs of RecQ helicase. (Yamagata, supra.) An indication of the existence and importance of mammalian repair systems is given by certain human hereditary disorders, including the recessive disease xeroderma pigmentosum (XP). This disorder results in a hypersensitivity to sunlight, especially ultraviolet, and produces skin defects. The recessive human disorder Bloom's syndrome results in an increased frequency of chromosomal aberrations, including sister chromosome exchanges. Bloom's syndrome suggests mammalian cells may have recombination-repair systems, since the BLN gene represents a RecQ helicase homolog. (Yamagata, supra)

The discovery of new DNA replication- and repair-associated proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of developmental disorders, proliferative disorders including cancer, and autoimmune/inflammatory disorders.

SUMMARY OF THE INVENTION

The invention features substantially purified polypeptides, DNA replication- and repair-associated proteins, referred to collectively as "DRASP". In one aspect, the invention provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, and SEQ ID NO:5 (SEQ ID NO:1-5), and fragments thereof.

The invention further provides a substantially purified variant having at least 90% amino acid identity to at least one of the amino acid sequences selected from the group consisting of SEQ ID NO:1-5, and fragments thereof. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and fragments thereof. The invention also includes an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and fragments thereof.

Additionally, the invention provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and fragments thereof.

The invention also provides an isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, and SEQ ID NO:10 (SEQ ID NO:6-10), and fragments thereof. The invention further provides an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, and fragments thereof.

The invention also provides a method for detecting a polynucleotide in a sample containing

nucleic acids, the method comprising the steps of (a) hybridizing the complement of the polynucleotide sequence to at least one of the polynucleotides of the sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide in the sample. In one aspect, the method further comprises amplifying the polynucleotide prior to hybridization.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and fragments thereof. In another aspect, the expression vector is contained within a host cell.

The invention also provides a method for producing a polypeptide, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a fragment of a polynucleotide under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide selected from the group consisting of SEQ ID NO:1-5, and fragments thereof. The invention also provides a purified agonist and a purified antagonist to the polypeptide.

The invention also provides a method for treating or preventing a disorder associated with decreased expression or activity of DRASP, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention also provides a method for treating or preventing a disorder associated with increased expression or activity of DRASP, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and fragments thereof.

BRIEF DESCRIPTION OF THE TABLES

Table 1 shows nucleotide and polypeptide sequence identification numbers (SEQ ID NO), clone identification numbers (clone ID), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding DRASP.

Table 2 shows features of each polypeptide sequence including potential motifs, homologous

sequences, and methods and algorithms used for identification of DRASP.

Table 3 shows the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis, diseases or disorders associated with these tissues, and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which Incyte cDNA clones encoding DRASP were isolated.

Table 5 shows the programs, their descriptions, references, and threshold parameters used to analyze DRASP.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"DRASP" refers to the amino acid sequences of substantially purified DRASP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which, when bound to DRASP, increases or prolongs

the duration of the effect of DRASP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of DRASP.

An "allelic variant" is an alternative form of the gene encoding DRASP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding DRASP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same as DRASP or a polypeptide with at least one functional characteristic of DRASP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding DRASP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding DRASP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent DRASP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of DRASP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms "amino acid" or "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments," "immunogenic fragments," or "antigenic fragments" refer to fragments of DRASP which are preferably at least 5 to about 15 amino acids in length, most preferably at least 14 amino acids, and which retain some biological activity or immunological activity of DRASP. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known

in the art.

The term "antagonist" refers to a molecule which, when bound to DRASP, decreases the amount or the duration of the effect of the biological or immunological activity of DRASP. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of DRASP.

The term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind DRASP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

The term "biologically active," refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic DRASP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" bonds to the complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid

strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" or a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding DRASP or fragments of DRASP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence"refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using XL-PCR kit (Perkin-Elmer, Norwalk CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW Fragment Assembly system (GCG, Madison WI). Some sequences have been both extended and assembled to produce the consensus sequence.

The term "correlates with expression of a polynucleotide" indicates that the detection of the presence of nucleic acids, the same or related to a nucleic acid sequence encoding DRASP, by northern analysis is indicative of the presence of nucleic acids encoding DRASP in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding DRASP.

A "deletion"refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

The term "similarity" refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the

completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" or "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MEGALIGN program (DNASTAR, Madison WI). The MEGALIGN program can create alignments between two or more sequences according to different methods, e.g., the clustal method. (See, e.g., Higgins, D.G. and P.M. Sharp (1988) Gene 73:237-244.) The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) Methods Enzymol. 183:626-645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance.

The term "humanized antibody" refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

The term "hybridization complex" refers to a complex formed between two nucleic acid

sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" or "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term "microarray" refers to an arrangement of distinct polynucleotides on a substrate.

The terms "element" or "array element" in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

The term "modulate" refers to a change in the activity of DRASP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of DRASP.

The phrases "nucleic acid" or "nucleic acid sequence," as used herein, refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material. In this context, "fragments" refers to those nucleic acid sequences which comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:6-10, for example, as distinct from any other sequence in the same genome. For example, a fragment of SEQ ID NO:6-10 is useful in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:6-10 from related polynucleotide sequences. A fragment of SEQ ID NO:6-10 is at least about 15-20 nucleotides in length. The precise length of the fragment of SEQ ID NO:6-10 and the region of SEQ ID NO:6-10 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment. In some cases, a fragment, when translated, would produce polypeptides retaining some functional characteristic, e.g., antigenicity, or structural domain characteristic, e.g., ATP-binding site, of the full-length polypeptide.

The terms "operably associated" or "operably linked" refer to functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the

promoter controls the translation of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the sequence encoding the polypeptide but still bind to operator sequences that control expression of the polypeptide.

The term "oligonucleotide" refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. "Oligonucleotide" is substantially equivalent to the terms "amplimer," "primer," "oligomer," and "probe," as these terms are commonly defined in the art.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding DRASP, or fragments thereof, or DRASP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" or "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "stringent conditions" refers to conditions which permit hybridization between polynucleotides and the claimed polynucleotides. Stringent conditions can be defined by salt concentration, the concentration of organic solvent, e.g., formamide, temperature, and other conditions well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acids or nucleotides by

different amino acids or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

"Transformation" describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "variant" of DRASP polypeptides refers to an amino acid sequence that is altered by one or more amino acid residues. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine). More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, LASERGENE software (DNASTAR).

The term "variant," when used in the context of a polynucleotide sequence, may encompass a polynucleotide sequence related to DRASP. This definition may also include, for example, "allelic" (as defined above), "splice," "species," or "polymorphic" variants. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or an absence of domains. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

THE INVENTION

The invention is based on the discovery of new human DNA replication- and repair-associated proteins (DRASP), the polynucleotides encoding DRASP, and the use of these compositions for the diagnosis, treatment, or prevention of developmental disorders, proliferative disorders including cancer, and autoimmune/inflammatory disorders.

Table 1 lists the Incyte Clones used to derive full length nucleotide sequences encoding DRASP. Columns 1 and 2 show the sequence identification numbers (SEQ ID NO) of the amino acid and nucleic acid sequences, respectively. Column 3 shows the Clone ID of the Incyte Clone in which nucleic acids encoding each DRASP were identified, and column 4, the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones, their corresponding cDNA libraries, and shotgun sequences useful as fragments in hybridization technologies, and which are part of the consensus nucleotide sequence of each DRASP.

The columns of Table 2 show various properties of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3, potential phosphorylation sites; column 4, potential glycosylation sites; column 5, the amino acid residues comprising signature sequences and motifs; column 6, the identity of each protein; and column 7, analytical methods used to identify each protein through sequence homology and protein motifs. The columns of Table 3 show the tissue-specificity and disease-association of nucleotide sequences encoding DRASP. The first column of Table 3 lists the polynucleotide sequence identifiers. The second column lists tissue categories which express DRASP as a fraction of total tissue categories expressing DRASP. The third column lists the disease classes associated with those tissues expressing DRASP. The fourth column lists the vectors used to subclone the cDNA library.

The following fragments of the nucleotide sequences encoding DRASP are useful in hybridization or amplification technologies to identify SEQ ID NO:6-10 and to distinguish between SEQ ID NO:6-10 and similar polynucleotide sequences. The useful fragments are the fragment of SEQ ID NO:6 from about nucleotide 90 to about nucleotide 120 (encoding the fragment of DRASP-1 from about amino acid 30 to about amino acid 40), the fragment of SEQ ID NO:7 from about nucleotide 165 to about nucleotide 195 (encoding the fragment of DRASP-2 from about amino acid 55 to about amino acid 65), the fragment of SEQ ID NO:8 from about nucleotide 833 to about nucleotide 889; the fragment of SEQ ID NO:9 from about nucleotide 848 to about nucleotide 892; and the fragment of SEQ ID NO:10 from about nucleotide 101 to about nucleotide 130.

The invention also encompasses DRASP variants. A preferred DRASP variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the DRASP amino acid sequence, and which contains at least one

functional or structural characteristic of DRASP.

The invention also encompasses polynucleotides which encode DRASP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:6-10, which encodes DRASP.

The invention also encompasses a variant of a polynucleotide sequence encoding DRASP. In particular, such a variant polynucleotide sequence will have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding DRASP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:6-10 which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:6-10. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of DRASP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding DRASP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring DRASP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode DRASP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring DRASP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding DRASP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding DRASP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode DRASP and DRASP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce

mutations into a sequence encoding DRASP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:6-10 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30°C, more preferably of at least about 37°C, and most preferably of at least about 42°C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30°C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37°C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 μg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42°C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50 % formamide, and 200 μg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps which follow hybridization can also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperature of at least about 25°C, more preferably of at least about 42°C, and most preferably of at least about 68°C. In a preferred embodiment, wash steps will occur at 25°C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art.

Methods for DNA sequencing are well known in the art and may be used to practice any of the

embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Perkin-Elmer), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the Hamilton MICROLAB 2200 (Hamilton, Reno NV), Peltier Thermal Cycler 200 (PTC200; MJ Research, Watertown MA) and the ABI CATALYST 800 (Perkin-Elmer). Sequencing is then carried out using either ABI 373 or 377 DNA Sequencing Systems (Perkin-Elmer) or the MEGABACE capillary electrophoresis system (Molecular Dynamics, Sunnyvale CA). The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding DRASP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-306). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been

size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Perkin-Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode DRASP may be cloned in recombinant DNA molecules that direct expression of DRASP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express DRASP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter DRASP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In another embodiment, sequences encoding DRASP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, and Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232.) Alternatively, DRASP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A Peptide Synthesizer (Perkin-Elmer). Additionally, the amino acid sequence of DRASP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g, Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY.)

In order to express a biologically active DRASP, the nucleotide sequences encoding DRASP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding DRASP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding DRASP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding DRASP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding DRASP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding DRASP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding DRASP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding DRASP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or pSPORT1 plasmid (Life Technologies). Ligation of sequences encoding DRASP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of DRASP are needed, e.g. for the production of antibodies, vectors which direct high level expression of DRASP may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of DRASP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Grant et al. (1987) Methods Enzymol. 153:516-54; and Scorer, C. A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of DRASP. Transcription of sequences encoding DRASP may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV. (Takamatsu, N. (1987) EMBO J. 6:307-311.) Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding DRASP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses DRASP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based

vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of DRASP in cell lines is preferred. For example, sequences encoding DRASP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*$^-$ or *apr*$^-$ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* or *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding DRASP is inserted within a marker gene sequence, transformed cells containing sequences encoding DRASP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding DRASP under the control of a single

promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding DRASP and that express DRASP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of DRASP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on DRASP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul MN, Sect. IV; Coligan, J. E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding DRASP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding DRASP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding DRASP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode DRASP may be designed to contain signal sequences which direct secretion of DRASP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Manassas, VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding DRASP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric DRASP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of DRASP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the DRASP encoding sequence and the heterologous protein sequence, so that DRASP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled DRASP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract systems (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, preferably ^{35}S-methionine.

Fragments of DRASP may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be

achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin-Elmer). Various fragments of DRASP may be synthesized separately and then combined to produce the full length molecule.

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between DRASP and DNA replication and repair proteins. In addition, the expression of DRASP is closely associated with cell proliferation, cancer, and inflammation. Therefore, DRASP appears to be associated with developmental disorders, proliferative disorders including cancer, and autoimmune/inflammatory disorders. In disorders where the expression or activity of DRASP is low, it is desirable to increase the expression or activity of DRASP. In disorders where the expression or activity of DRASP is high, it is desirable to decrease the expression or activity of DRASP.

Therefore, in one embodiment, DRASP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder in which the expression or activity of DRASP is decreased. Examples of such a disorder include, but are not limited to, a developmental disorder such as renal tubular acidosis, anemia, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, and cataract; a proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia (including Bloom's syndrome), lymphoma, melanoma, myeloma, multiple endocrine neoplasia (Werner's syndrome), sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic

anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma.

In another embodiment, a vector capable of expressing DRASP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified DRASP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of DRASP may be administered to a subject to treat or prevent a disorder including, but not limited to, those listed above.

In a further embodiment, an antagonist of DRASP may be administered to a subject to treat or prevent a disorder in which the expression or activity of DRASP is increased. Examples of such a disorder include, but are not limited to, those described above. In one aspect, an antibody which specifically binds DRASP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express DRASP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding DRASP may be administered to a subject to treat or prevent a disorder including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of DRASP may be produced using methods which are generally known in the art. In particular, purified DRASP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind DRASP. Antibodies to DRASP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and

others may be immunized by injection with DRASP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to DRASP have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of DRASP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to DRASP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce DRASP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D.R. (1991) Proc. Natl. Acad. Sci. 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for DRASP may also be generated. For example, such fragments include, but are not limited to, F(ab')2 fragments produced by pepsin

digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between DRASP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering DRASP epitopes is preferred, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for DRASP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of DRASP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple DRASP epitopes, represents the average affinity, or avidity, of the antibodies for DRASP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular DRASP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the DRASP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of DRASP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington, D. C.; and Liddell, J. E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York, NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is preferred for use in procedures requiring precipitation of DRASP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding DRASP, or any

fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding DRASP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding DRASP. Thus, complementary molecules or fragments may be used to modulate DRASP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding DRASP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding DRASP. (See, e.g., Sambrook, supra; Ausubel, 1995, supra.)

Genes encoding DRASP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding DRASP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding DRASP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze

endonucleolytic cleavage of sequences encoding DRASP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding DRASP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nature Biotechnology 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the

therapeutic effects discussed above. Such pharmaceutical compositions may consist of DRASP, antibodies to DRASP, and mimetics, agonists, antagonists, or inhibitors of DRASP. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of

gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of DRASP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An

animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example DRASP or fragments thereof, antibodies of DRASP, and agonists, antagonists or inhibitors of DRASP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of therapeutic to toxic effects is the therapeutic index, and it can be expressed as the ED_{50}/LD_{50} ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind DRASP may be used for the diagnosis of disorders characterized by expression of DRASP, or in assays to monitor patients being treated with DRASP or agonists, antagonists, or inhibitors of DRASP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for DRASP include methods which utilize the antibody and a label to detect DRASP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and

may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring DRASP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of DRASP expression. Normal or standard values for DRASP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to DRASP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of DRASP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding DRASP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of DRASP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of DRASP, and to monitor regulation of DRASP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding DRASP or closely related molecules may be used to identify nucleic acid sequences which encode DRASP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding DRASP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably have at least 50% sequence identity to any of the DRASP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:6-10 or from genomic sequences including promoters, enhancers, and introns of the DRASP gene.

Means for producing specific hybridization probes for DNAs encoding DRASP include the cloning of polynucleotide sequences encoding DRASP or DRASP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding DRASP may be used for the diagnosis of a disorder associated with expression of DRASP. Examples of such a disorder include, but are not limited to, a developmental disorder such as renal tubular acidosis, anemia, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, and cataract; a proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia (including Bloom's syndrome), lymphoma, melanoma, myeloma, multiple endocrine neoplasia (Werner's syndrome), sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma. The polynucleotide sequences encoding DRASP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and ELISA assays; and in microarrays utilizing fluids or tissues from patients to detect altered DRASP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding DRASP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding DRASP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable

incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding DRASP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of DRASP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding DRASP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding DRASP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding DRASP, or a fragment of a polynucleotide complementary to the polynucleotide encoding DRASP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of DRASP include radiolabeling

or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding DRASP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding DRASP on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may

reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, DRASP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between DRASP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with DRASP, or fragments thereof, and washed. Bound DRASP is then detected by methods well known in the art. Purified DRASP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding DRASP specifically compete with a test compound for binding DRASP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with DRASP.

In additional embodiments, the nucleotide sequences which encode DRASP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. [Atty Docket No. PF-0563 P, filed August 7, 1998] and U.S. Ser. No. [Atty Docket No. PF-0580 P, filed August 27, 1998], are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Valencia CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6). Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), pSPORT1 plasmid (Life Technologies), or pINCY (Incyte Pharmaceuticals, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids were recovered from host cells by in vivo excision, using the UNIZAP vector system (Stratagene) or cell lysis. Plasmids were purified using at least one of the following: a Magic or

WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the REAL Prep 96 plasmid kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a Fluoroskan II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

The cDNAs were prepared for sequencing using the ABI CATALYST 800 (Perkin-Elmer) or the HYDRA microdispenser (Robbins Scientific, Sunnyvale CA) or MICROLAB 2200 (Hamilton, Reno NV) systems in combination with the PTC-200 thermal cyclers (MJ Research, Watertown MA). The cDNAs were sequenced using the ABI PRISM 373 or 377 sequencing systems (Perkin-Elmer) and standard ABI protocols, base calling software, and kits. In one alternative, cDNAs were sequenced using the MEGABACE 1000 capillary electrophoresis sequencing system (Molecular Dynamics, Sunnyvale CA). In another alternative, the cDNAs were amplified and sequenced using the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer). In yet another alternative, cDNAs were sequenced using solutions and dyes from Amersham Pharmacia Biotech. Reading frames for the ESTs were determined using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA, extension, and shotgun sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the software programs, descriptions, references, and threshold parameters used. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides a brief description thereof, the third column presents the references which are incorporated by reference herein, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the probability the greater the homology). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, S. San Francisco CA) and LASERGENE software (DNASTAR).

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS to acquire annotation, using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, PFAM, and Prosite.

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:8-14. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ database (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analyses are reported a percentage distribution of libraries in which the transcript encoding DRASP occurred. Analysis involved the categorization of cDNA libraries by

organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease categories included cancer, inflammation/trauma, fetal, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease expression are reported in Table 3.

V. Extension of DRASP Encoding Polynucleotides

Full length nucleic acid sequences (SEQ ID NO:6 and 7) were produced by extension of the clones described in the Invention section, using oligonucleotide primers based on those fragments. One primer was synthesized to initiate extension of an antisense polynucleotide, and the other was synthesized to initiate extension of a sense polynucleotide. Primers were used to facilitate the extension of the known sequence "outward" generating amplicons containing new unknown nucleotide sequence for the region of interest. The initial primers were designed from the cDNA using OLIGO 4.06 software (National Biosciences, Plymouth, MN), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries (Life Technologies) were used to extend the sequence. If more than one extension is necessary or desired, additional sets of primers are designed to further extend the known region.

High fidelity amplification was obtained by following the instructions for the XL-PCR kit (The Perkin-Elmer Corp., Norwalk, CT) and thoroughly mixing the enzyme and reaction mix. PCR was performed using the PTC-200 thermal cycler (MJ Research, Inc., Watertown, MA), beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, with the following parameters:

Step 1	94° C for 1 min (initial denaturation)
Step 2	65° C for 1 min
Step 3	68° C for 6 min
Step 4	94° C for 15 sec
Step 5	65° C for 1 min
Step 6	68° C for 7 min
Step 7	Repeat steps 4 through 6 for an additional 15 cycles
Step 8	94° C for 15 sec
Step 9	65° C for 1 min
Step 10	68° C for 7:15 min
Step 11	Repeat steps 8 through 10 for an additional 12 cycles
Step 12	72° C for 8 min
Step 13	4° C (and holding)

A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6% to 0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were excised from the gel, purified using QIAQUICK purification kit (Qiagen, Inc.), and trimmed of overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13 μl of ligation buffer, 1μl T4-DNA ligase (15 units) and 1μl T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2 to 3 hours, or overnight at 16° C. Competent E. coli cells (in 40 μl of appropriate media) were transformed with 3 μl of ligation mixture and cultured in 80 μl of SOC medium. (See, e.g., Sambrook, supra, Appendix A, p. 2.) After incubation for one hour at 37°C, the E. coli mixture was plated on Luria Bertani (LB) agar (See, e.g., Sambrook, supra, Appendix A, p. 1) containing carbenicillin (2x carb). The following day, several colonies were randomly picked from each plate and cultured in 150 μl of liquid LB/2x carb medium placed in an individual well of an appropriate commercially-available sterile 96-well microtiter plate. The following day, 5 μl of each overnight culture was transferred into a non-sterile 96-well plate and, after dilution 1:10 with water, 5 μl from each sample was transferred into a PCR array.

For PCR amplification, 18 μl of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions:

Step 1 94° C for 60 sec
Step 2 94° C for 20 sec
Step 3 55° C for 30 sec
Step 4 72° C for 90 sec
Step 5 Repeat steps 2 through 4 for an additional 29 cycles
Step 6 72° C for 180 sec
Step 7 4° C (and holding)

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid, and sequenced.

In like manner, the nucleotide sequences of SEQ ID NO:6 and SEQ ID NO:7 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for 5' extension, and an appropriate genomic library.

The full length nucleic acid sequence of SEQ ID NO:8-10 was produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other

primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICO GREEN quantitation reagent (0.25% (v/v) PICO GREEN) (Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly, MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulphoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer).

In like manner, the nucleotide sequence of SEQ ID NO:8-10 is used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for such extension, and an appropriate genomic library.

VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:6-10 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba1, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT-AR film (Eastman Kodak, Rochester NY) is exposed to the blots to film for several hours, hybridization patterns are compared visually.

VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV,

chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

VIII. Complementary Polynucleotides

Sequences complementary to the DRASP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring DRASP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of DRASP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the DRASP-encoding transcript.

IX. Expression of DRASP

Expression and purification of DRASP is achieved using bacterial or virus-based expression systems. For expression of DRASP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express DRASP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of DRASP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus

(AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding DRASP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E. K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, DRASP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from DRASP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch 10 and 16). Purified DRASP obtained by these methods can be used directly in the following activity assay.

X. Demonstration of DRASP Activity

DRASP activity is measured as incorporation of [^{32}P]dATP into a plasmid treated with a DNA damaging agent, such as cisplatin or ultraviolet irradiation, relative to a control, untreated plasmid DNA. (F. Coudore, et al. (1997) FEBS Lett. 414:581-584). Cell extracts are purified from mammalian cell lines, E. coli, or S. cerevisiae having compromised endogenous repair activities due to mutations in repair enzymes. Cell extracts are prepared by hypotonic lysis of cells followed by centrifugation at 300,000xg. Extracts are treated with 63% ammonium sulfate to minimize non-specific nuclease activity. The repair synthesis assay is performed in a 50 µl reaction volume containing 200 µg protein in cell extract, 300 ng damaged plasmid, 300 ng control plasmid, 4 µM dATP, 20 µM each dCTP, dTTP, and dGTP, 0.2 µM [^{32}P]dATP, 20 mM HEPES-KOH (pH 7.8), 2.5 µg creatine phosphokinase, 7 mM $MgCl_2$, and 2 mM EGTA. Identical reactions were set up with and without purified DRASP. After a 3 h incubation at 30°C, reaction mixtures are treated with 200 µg/ml proteinase K and 0.5% SDS. Plasmid DNA is purified from reaction mixtures by phenol-chloroform extraction and ethanol precipitation. Data is quantified by gel electrophoresis of linearized plasmid followed by

autoradiography, scintillation counting of excised DNA bands, and densitometry of the photographic negative of the gel to normalize for plasmid DNA recovery.

Alternatively, DRASP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled DRASP, washed, and any wells with labeled DRASP complex are assayed. Data obtained using different concentrations of DRASP are used to calculate values for the number, affinity, and association of DRASP with the candidate molecules.

XI. Functional Assays

DRASP function is assessed by expressing the sequences encoding DRASP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP, and to evaluate properties, for example, their apoptotic state. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of DRASP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding DRASP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells

using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding DRASP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XII. Production of DRASP Specific Antibodies

DRASP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the DRASP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides 15 residues in length are synthesized using an ABI 431A Peptide Synthesizer (Perkin-Elmer) using fmoc-chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity by, for example, binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIII. Purification of Naturally Occurring DRASP Using Specific Antibodies

Naturally occurring or recombinant DRASP is substantially purified by immunoaffinity chromatography using antibodies specific for DRASP. An immunoaffinity column is constructed by covalently coupling anti-DRASP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing DRASP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of DRASP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/DRASP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and DRASP is collected.

XIV. Identification of Molecules Which Interact with DRASP

DRASP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate molecules previously arrayed in the

wells of a multi-well plate are incubated with the labeled DRASP, washed, and any wells with labeled DRASP complex are assayed. Data obtained using different concentrations of DRASP are used to calculate values for the number, affinity, and association of DRASP with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	6	1817927	PROSNOT20	1817927H1 (PROSNOT20), 923538H1 (RATRNOT02), 2211972F6 (SINTFET03), 1574064F1 (LNODNOT03), and 1980391R6 (LUNGTUT03)
2	7	2678913	SINIUCT01	2678913H1 (SINIUCT01), 2828136H1 (TLYMNOT03), 481871R6 (HNT2RAT01), 3035287H1 (TLYMNOT05), 1330623F1 (PANCNOT07), 1312827F1 (BLADTUT02), and 190030F1 (SYNORAB01)
3	8	1849911	LUNGFET03	159979R1 (ADENINB01), 222584R1 (PANCNOT01), 1290029F1 and 1290029F6 (BRAINOT11), 1485328F1 (CORPNOT02), 1553837H1 (BLADTUT04), 1575237F1 (LNODNOT03), 1849911F6 and 1849911H1 (LUNGFET03)
4	9	2223353	LUNGNOT18	1295378F1 (PGANNOT03), 2048588H1 (LIVRFET02), 2223353H1 (LUNGNOT18), 2736221F6 (OVARNOT09), 3332938F6 (BRAIFET01)
5	10	2593035	LUNGNOT22	1984482R6 (LUNGAST01), 1697341T6 (COLNNOT23), 1633837F6 (COLNNOT19), 2593035H1 (LUNGNOT22)

Table 2

Protein Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
1	211	T5 S117 S131		N-Myristoylation Sites: G12 G143 Annexin type I: S111-I127 I130-D138	replication-associated protein	MOTIFS PRINTS
2	353	T81 S141 S146 S323 S324 S42 S59 T70 S106 S136 S141 S145 S146 S172 S255 S302 T307	N86 N158 N298	RecF: Q235-K252	XPMC2 protein (prevents mitotic catastrophe in yeast)	BLAST BLOCKS
3	746	T351 S24 T25 T26 S127 S136 S187 S215 S318 T325 T373 T570 S571 S591 T617 S676 S685 T707 T717 T732 S264 T422 T566 T570 T643 Y653 Y698	N165 N234 N565	I323-V340	DNA recombination and repair protein	PRINTS
4	346	T39 S23 S31 S42 T189 T241 T314	N40	G107-T114	Rad51-related protein	MOTIFS BLAST
5	194	S121 T163 S137 S176			mutagenic suppressor HAM1-related	BLAST

Table 3

Polynucleotide Seq ID NO:	Tissue Expression (Fraction of Total)	Disease Class (Fraction of Total)	Vector
6	Reproductive (0.27) Hematopoietic/Immune (0.16) Nervous (0.15)	Immortalized, Cancerous, or Proliferating (0.65) Inflammation (0.36)	pINCY
7	Reproductive (0.26) Cardiovascular (0.17) Hematopoietic/Immune (0.17) Developmental (0.13)	Immortalized, Cancerous, or Proliferating (0.65) Inflammation (0.31)	pINCY
8	Reproductive (0.232) Gastrointestinal (0.176) Nervous (0.176) Hematopoietic/Immune (0.148)	Cancer/Proliferative (0.641) Inflammation (0.246)	pINCY
9	Gastrointestinal (0.182) Nervous (0.182) Reproductive (0.182)	Cancer/Proliferative (0.636) Inflammation (0.182)	pINCY
10	Reproductive (0.224) Hematopoietic/Immune (0.163) Nervous (0.153) Gastrointestinal (0.133)	Cancer/Proliferative (0.673) Inflammation (0.245)	pINCY

Table 4

Polynucleotide SEQ ID NO:	Clone ID	Library	Library Comment
6	1817927	PROSNOT20	Library was constructed using RNA isolated from diseased prostate tissue removed from a 65-year-old Caucasian male during a radical prostatectomy. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma.
7	2678913	SINIUCT01	Library was constructed using RNA isolated from ileum tissue obtained from a 42-year-old Caucasian male during a total intra-abdominal colectomy and endoscopic jejunostomy. Previous surgeries included polypectomy, colonoscopy, and spinal canal exploration. Family history included cerebrovascular disease, benign hypertension, atherosclerotic coronary artery disease, and type II diabetes.
8	1849911	LUNGFET03	Library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
9	2223353	LUNGNOT18	Library was constructed using RNA isolated from left upper lobe lung tissue removed from a 66-year-old Caucasian female. Pathology for the associated tumor tissue indicated a grade 2 adenocarcinoma. Patient history included cerebrovascular and atherosclerotic coronary artery disease. Family history included a myocardial infarction and atherosclerotic coronary artery disease.
10	2593035	LUNGNOT22	Library was constructed using RNA isolated from lung tissue removed from a 58-year-old Caucasian female. The tissue sample used to construct this library was found to have tumor contaminant upon microscopic examination. Pathology for the associated tumor tissue indicated a caseating granuloma. Family history included congestive heart failure, breast cancer, secondary bone cancer, acute myocardial infarction and atherosclerotic coronary artery disease.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS and PRINTS databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and Probability value= 1.0E-3 or less
PFAM	A Hidden Markov Models-based application useful for protein family search.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits, depending on individual protein families

Table 5 cont.

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Score= 4.0 or greater
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. supra; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and SEQ ID NO:5 and fragments thereof.

2. A substantially purified variant having at least 90% amino acid sequence identity to the amino acid sequence of claim 1.

3. An isolated and purified polynucleotide encoding the polypeptide of claim 1.

4. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 3.

5. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide of claim 3.

6. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 3.

7. A method for detecting a polynucleotide, the method comprising the steps of:

(a) hybridizing the polynucleotide of claim 6 to at least one nucleic acid in a sample, thereby forming a hybridization complex; and

(b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the polynucleotide in the sample.

8. The method of claim 7 further comprising amplifying the polynucleotide prior to hybridization.

9. An isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10 and fragments thereof.

10. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 9.

11. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 9.

12. An expression vector comprising at least a fragment of the polynucleotide of claim 3.

13. A host cell comprising the expression vector of claim 12.

14. A method for producing a polypeptide, the method comprising the steps of:

a) culturing the host cell of claim 13 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

15. A pharmaceutical composition comprising the polypeptide of claim 1 in conjunction

with a suitable pharmaceutical carrier.

16. A purified antibody which specifically binds to the polypeptide of claim 1.

17. A purified agonist of the polypeptide of claim 1.

18. A purified antagonist of the polypeptide of claim 1.

19. A method for treating or preventing a disorder associated with decreased expression or activity of DRASP, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 15.

20. A method for treating or preventing a disorder associated with increased expression or activity of DRASP, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 18.

SEQUENCE LISTING

<110> INCYTE PHARMACEUTICALS, INC.
TANG, Y. Tom
CORLEY, Neil C.
BAUGHN, Mariah R.
REDDY, Roopa
GUEGLER, Karl J.
YUE, Henry

<120> DNA REPLICATION- AND REPAIR-ASSOCIATED PROTEINS

<130> PF-0563 PCT

<140> To Be Assigned
<141> Herewith

<150> 09/130,650; unassigned; 09/140,982; unassigned
<151> 1998-08-07; 1998-08-07; 1998-08-27; 1998-08-27

<160> 10

<170> PERL Program

<210> 1
<211> 211
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Clone No: 1817927

<400> 1

Met	Thr	Thr	Leu	Thr	Arg	Gln	Asp	Leu	Asn	Phe	Gly	Gln	Val	Val	
1				5				10						15	
Ala	Asp	Val	Leu	Cys	Glu	Phe	Leu	Glu	Val	Ala	Val	His	Leu	Ile	
				20				25						30	
Leu	Tyr	Val	Arg	Glu	Val	Tyr	Pro	Val	Gly	Ile	Phe	Gln	Lys	Arg	
				35				40						45	
Lys	Lys	Tyr	Asn	Val	Pro	Val	Gln	Met	Ser	Cys	His	Pro	Glu	Leu	
				50				55						60	
Asn	Gln	Tyr	Ile	Gln	Asp	Thr	Leu	His	Cys	Val	Lys	Pro	Leu	Leu	
				65				70						75	
Glu	Lys	Asn	Asp	Val	Glu	Lys	Val	Val	Val	Val	Ile	Leu	Asp	Lys	
				80				85						90	
Glu	His	Arg	Pro	Val	Glu	Lys	Phe	Val	Phe	Glu	Ile	Thr	Gln	Pro	
				95				100						105	
Pro	Leu	Leu	Ser	Ile	Ser	Ser	Asp	Ser	Leu	Leu	Ser	His	Val	Glu	
				110				115						120	
Gln	Leu	Leu	Arg	Ala	Phe	Ile	Leu	Lys	Ile	Ser	Val	Cys	Asp	Ala	
				125				130						135	
Val	Leu	Asp	His	Asn	Pro	Pro	Gly	Cys	Thr	Phe	Thr	Val	Leu	Val	
				140				145						150	
His	Thr	Arg	Glu	Ala	Ala	Thr	Arg	Asn	Met	Glu	Lys	Ile	Gln	Val	
				155				160						165	

```
Ile Lys Asp Phe Pro Trp Ile Leu Ala Asp Glu Gln Asp Val His
                170                 175                 180
Met His Asp Pro Arg Leu Ile Pro Leu Lys Thr Met Thr Ser Asp
                185                 190                 195
Ile Leu Lys Met Gln Leu Tyr Val Glu Glu Arg Ala His Lys Gly
                200                 205                 210
Ser
```

<210> 2
<211> 353
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Clone No: 2678913

<400> 2

```
Met Ser Thr Leu Leu Leu Asn Leu Asp Phe Gly Glu Pro Pro Pro
  1               5                  10                  15
Lys Lys Ala Leu Glu Gly Asn Ala Lys His Arg Asn Phe Val Lys
                20                  25                  30
Lys Arg Arg Leu Leu Glu Arg Arg Gly Phe Leu Ser Lys Lys Asn
                35                  40                  45
Gln Pro Pro Ser Lys Ala Pro Lys Leu His Ser Glu Pro Ser Lys
                50                  55                  60
Lys Gly Glu Thr Pro Thr Val Asp Gly Thr Trp Lys Thr Pro Ser
                65                  70                  75
Phe Pro Lys Lys Lys Thr Ala Ala Ser Ser Asn Gly Ser Gly Gln
                80                  85                  90
Pro Leu Asp Lys Lys Ala Ala Val Ser Trp Leu Thr Pro Ala Pro
                95                 100                 105
Ser Lys Lys Ala Asp Ser Val Ala Ala Lys Val Asp Leu Leu Gly
                110                 115                 120
Glu Phe Gln Ser Ala Leu Pro Lys Ile Asn Ser His Pro Thr Arg
                125                 130                 135
Ser Gln Lys Lys Ser Ser Gln Lys Lys Ser Ser Lys Lys Asn His
                140                 145                 150
Pro Gln Lys Asn Ala Pro Gln Asn Ser Thr Gln Ala His Ser Glu
                155                 160                 165
Asn Lys Cys Ser Gly Ala Ser Gln Lys Leu Pro Arg Lys Met Val
                170                 175                 180
Ala Ile Asp Cys Glu Met Val Gly Thr Gly Pro Lys Gly His Val
                185                 190                 195
Ser Ser Leu Ala Arg Cys Ser Ile Val Asn Tyr Asn Gly Asp Val
                200                 205                 210
Leu Tyr Asp Glu Tyr Ile Leu Pro Pro Cys His Ile Val Asp Tyr
                215                 220                 225
Arg Thr Arg Trp Ser Gly Ile Arg Lys Gln His Met Val Asn Ala
                230                 235                 240
Thr Pro Phe Lys Ile Ala Arg Gly Gln Ile Leu Lys Ile Leu Thr
                245                 250                 255
Gly Lys Ile Val Val Gly His Ala Ile His Asn Asp Phe Lys Ala
                260                 265                 270
```

```
Leu Gln Tyr Phe His Pro Lys Ser Leu Thr Arg Asp Thr Ser His
                275                 280                 285
Ile Pro Pro Leu Asn Arg Lys Ala Asp Cys Pro Glu Asn Ala Thr
                290                 295                 300
Met Ser Leu Lys His Leu Thr Lys Lys Leu Leu Asn Arg Asp Ile
                305                 310                 315
Gln Val Gly Lys Ser Gly His Ser Ser Val Glu Asp Ala Gln Ala
                320                 325                 330
Thr Met Glu Leu Tyr Lys Leu Val Glu Val Glu Trp Glu Glu His
                335                 340                 345
Leu Ala Arg Asn Pro Pro Thr Asp
                350
```

<210> 3
<211> 746
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Clone No: 1849911

<400> 3

```
Met Arg Ile Val Tyr Arg Met Arg Gly Leu Leu Gly Asp Ala Thr
  1                 5                  10                  15
Glu Glu Phe Ile Glu Ser Leu Asp Ser Thr Thr Asp Glu Glu Glu
                 20                  25                  30
Asp Glu Glu Glu Val Tyr Lys Met Ala Gly Val Met Ala Gln Cys
                 35                  40                  45
Gly Gly Leu Glu Cys Met Leu Asn Arg Leu Ala Gly Ile Arg Asp
                 50                  55                  60
Phe Lys Gln Gly Arg His Leu Leu Thr Val Leu Leu Lys Leu Phe
                 65                  70                  75
Ser Tyr Cys Val Lys Val Lys Val Asn Arg Gln Gln Leu Val Lys
                 80                  85                  90
Leu Glu Met Asn Thr Leu Asn Val Met Leu Gly Thr Leu Asn Leu
                 95                 100                 105
Ala Leu Val Ala Glu Gln Glu Ser Lys Asp Ser Gly Gly Ala Ala
                110                 115                 120
Val Ala Glu Gln Val Leu Ser Ile Met Glu Ile Ile Leu Asp Glu
                125                 130                 135
Ser Asn Ala Glu Pro Leu Ser Glu Asp Lys Gly Asn Leu Leu Leu
                140                 145                 150
Thr Gly Asp Lys Asp Gln Leu Val Met Leu Leu Asp Gln Ile Asn
                155                 160                 165
Ser Thr Phe Val Arg Ser Asn Pro Ser Val Leu Gln Gly Leu Leu
                170                 175                 180
Arg Ile Ile Pro Tyr Leu Ser Phe Gly Glu Val Glu Lys Met Gln
                185                 190                 195
Ile Leu Val Glu Arg Phe Lys Pro Tyr Cys Asn Phe Asp Lys Tyr
                200                 205                 210
Asp Glu Asp His Ser Gly Asp Asp Lys Val Phe Leu Asp Cys Phe
                215                 220                 225
Cys Lys Ile Ala Ala Gly Ile Lys Asn Asn Ser Asn Gly His Gln
```

```
                230                 235                 240
Leu Lys Asp Leu Ile Leu Gln Lys Gly Ile Thr Gln Asn Ala Leu
                245                 250                 255
Asp Tyr Met Lys Lys His Ile Pro Ser Ala Lys Asn Leu Asp Ala
                260                 265                 270
Asp Ile Trp Lys Lys Phe Leu Ser Arg Pro Ala Leu Pro Phe Ile
                275                 280                 285
Leu Arg Leu Leu Arg Gly Leu Ala Ile Gln His Pro Gly Thr Gln
                290                 295                 300
Val Leu Ile Gly Thr Asp Ser Ile Pro Asn Leu His Lys Leu Glu
                305                 310                 315
Gln Val Ser Ser Asp Glu Gly Ile Gly Thr Leu Ala Glu Asn Leu
                320                 325                 330
Leu Glu Ala Leu Arg Glu His Pro Asp Val Asn Lys Lys Ile Asp
                335                 340                 345
Ala Ala Arg Arg Glu Thr Arg Ala Glu Lys Lys Arg Met Ala Met
                350                 355                 360
Ala Met Arg Gln Lys Ala Leu Gly Thr Leu Gly Met Thr Thr Asn
                365                 370                 375
Glu Lys Gly Gln Val Val Thr Lys Thr Ala Leu Leu Lys Gln Met
                380                 385                 390
Glu Glu Leu Ile Glu Glu Pro Gly Leu Thr Cys Cys Ile Cys Arg
                395                 400                 405
Glu Gly Tyr Lys Phe Gln Pro Thr Lys Val Leu Gly Ile Tyr Thr
                410                 415                 420
Phe Thr Lys Arg Val Ala Leu Glu Glu Leu Glu Asn Lys Pro Arg
                425                 430                 435
Lys Gln Gln Gly Tyr Ser Thr Val Ser His Phe Asn Ile Val His
                440                 445                 450
Tyr Asp Cys His Leu Ala Ala Val Arg Leu Ala Arg Gly Arg Glu
                455                 460                 465
Glu Trp Glu Ser Ala Ala Leu Gln Asn Ala Asn Thr Lys Cys Asn
                470                 475                 480
Ala Leu Leu Pro Val Trp Gly Pro His Val Pro Glu Ser Ala Phe
                485                 490                 495
Ala Thr Cys Leu Ala Arg His Asn Thr Tyr Leu Gln Glu Cys Thr
                500                 505                 510
Gly Gln Arg Glu Pro Thr Tyr Gln Leu Asn Ile His Asp Ile Lys
                515                 520                 525
Leu Leu Phe Leu Arg Phe Ala Met Glu Gln Ser Phe Ser Ala Asp
                530                 535                 540
Thr Gly Gly Gly Gly Arg Glu Ser Asn Ile His Leu Ile Pro Tyr
                545                 550                 555
Ile Ile His Thr Val Leu Tyr Val Leu Asn Thr Thr Arg Ala Thr
                560                 565                 570
Ser Arg Glu Glu Lys Asn Leu Gln Gly Phe Leu Glu Gln Pro Lys
                575                 580                 585
Glu Lys Trp Val Glu Ser Ala Phe Glu Val Asp Gly Pro Tyr Tyr
                590                 595                 600
Phe Thr Val Leu Ala Leu His Ile Leu Pro Pro Glu Gln Trp Arg
                605                 610                 615
Ala Thr Arg Val Glu Ile Leu Arg Arg Leu Leu Val Thr Ser Gln
                620                 625                 630
Ala Arg Ala Val Ala Pro Gly Gly Ala Thr Arg Leu Thr Asp Lys
                635                 640                 645
Ala Val Lys Asp Tyr Ser Ala Tyr Arg Ser Ser Leu Leu Phe Trp
                650                 655                 660
```

```
Ala Leu Val Asp Leu Ile Tyr Asn Met Phe Lys Lys Val Pro Thr
                665                 670                 675
Ser Asn Thr Glu Gly Gly Trp Ser Cys Ser Leu Ala Glu Tyr Ile
                680                 685                 690
Arg His Asn Asp Met Pro Ile Tyr Glu Ala Ala Asp Lys Ala Leu
                695                 700                 705
Lys Thr Phe Gln Glu Glu Phe Met Pro Val Glu Thr Phe Ser Glu
                710                 715                 720
Phe Leu Asp Val Ala Gly Leu Leu Ser Glu Ile Thr Asp Pro Glu
                725                 730                 735
Ser Phe Leu Lys Asp Leu Leu Asn Ser Val Pro
                740                 745
```

<210> 4
<211> 346
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Clone No: 2223353

<400> 4

```
Met Asp Leu Asp Leu Leu Asp Leu Asn Pro Arg Ile Ile Ala Ala
  1               5                  10                  15
Ile Lys Lys Ala Lys Leu Lys Ser Val Lys Glu Val Leu His Phe
                20                  25                  30
Ser Gly Pro Asp Leu Lys Arg Leu Thr Asn Leu Ser Ser Pro Glu
                35                  40                  45
Val Trp His Leu Leu Arg Thr Ala Ser Leu His Leu Arg Gly Ser
                50                  55                  60
Ser Ile Leu Thr Ala Leu Gln Leu His Gln Gln Lys Glu Arg Phe
                65                  70                  75
Pro Thr Gln His Gln Arg Leu Ser Leu Gly Cys Pro Val Leu Asp
                80                  85                  90
Ala Leu Leu Arg Gly Gly Leu Pro Leu Asp Gly Ile Thr Glu Leu
                95                 100                 105
Ala Gly Arg Ser Ser Ala Gly Lys Thr Gln Leu Ala Leu Gln Leu
               110                 115                 120
Cys Leu Ala Val Gln Phe Pro Arg Gln His Gly Gly Leu Glu Ala
               125                 130                 135
Gly Ala Val Tyr Ile Cys Thr Glu Asp Ala Phe Pro His Lys Arg
               140                 145                 150
Leu Gln Gln Leu Met Ala Gln Gln Pro Arg Leu Arg Thr Asp Val
               155                 160                 165
Pro Gly Glu Leu Leu Gln Lys Leu Arg Phe Gly Ser Gln Ile Phe
               170                 175                 180
Ile Glu His Val Ala Asp Val Asp Thr Leu Leu Glu Cys Val Asn
               185                 190                 195
Lys Lys Val Pro Val Leu Leu Ser Arg Gly Met Ala Arg Leu Val
               200                 205                 210
Val Ile Asp Ser Val Ala Ala Pro Phe Arg Cys Glu Phe Asp Ser
               215                 220                 225
Gln Ala Ser Ala Pro Arg Ala Arg His Leu Gln Ser Leu Gly Ala
```

```
                230                 235                 240
Thr Leu Arg Glu Leu Ser Ser Ala Phe Gln Ser Pro Val Leu Cys
                245                 250                 255
Ile Asn Gln Val Thr Glu Ala Met Glu Glu Gln Gly Ala Ala His
                260                 265                 270
Gly Pro Leu Gly Phe Trp Asp Glu Arg Val Ser Pro Ala Leu Gly
                275                 280                 285
Ile Thr Trp Ala Asn Gln Leu Leu Val Arg Leu Leu Ala Asp Arg
                290                 295                 300
Leu Arg Glu Glu Glu Ala Ala Leu Gly Cys Pro Ala Arg Thr Leu
                305                 310                 315
Arg Val Leu Ser Ala Pro His Leu Pro Pro Ser Ser Cys Ser Tyr
                320                 325                 330
Thr Ile Ser Ala Glu Gly Val Arg Gly Thr Pro Gly Thr Gln Ser
                335                 340                 345
His
```

<210> 5
<211> 194
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Clone No: 2593035

<400> 5

```
Met Ala Ala Ser Leu Val Gly Lys Lys Ile Val Phe Val Thr Gly
  1               5                  10                  15
Asn Ala Lys Lys Leu Glu Glu Val Val Gln Ile Leu Gly Asp Lys
                 20                  25                  30
Phe Pro Cys Thr Leu Val Ala Gln Lys Ile Asp Leu Pro Glu Tyr
                 35                  40                  45
Gln Gly Glu Pro Asp Glu Ile Ser Ile Gln Lys Cys Gln Glu Ala
                 50                  55                  60
Val Arg Gln Val Gln Gly Pro Val Leu Val Glu Asp Thr Cys Leu
                 65                  70                  75
Cys Phe Asn Ala Leu Gly Gly Leu Pro Gly Pro Tyr Ile Lys Trp
                 80                  85                  90
Phe Leu Glu Lys Leu Lys Pro Glu Gly Leu His Gln Leu Leu Ala
                 95                 100                 105
Gly Phe Glu Asp Lys Ser Ala Tyr Ala Leu Cys Thr Phe Ala Leu
                110                 115                 120
Ser Thr Gly Asp Pro Ser Gln Pro Val Arg Leu Phe Arg Gly Arg
                125                 130                 135
Thr Ser Gly Arg Ile Val Ala Pro Arg Gly Cys Gln Asp Phe Gly
                140                 145                 150
Trp Asp Pro Cys Phe Gln Pro Asp Gly Tyr Glu Gln Thr Tyr Ala
                155                 160                 165
Glu Met Pro Lys Ala Glu Lys Asn Ala Val Ser His Arg Phe Arg
                170                 175                 180
Ala Leu Leu Glu Leu Gln Glu Tyr Phe Gly Ser Leu Ala Ala
                185                 190
```

<210> 6
<211> 1052
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Clone No: 1817927

<400> 6
cccgggttga gcgggaggcg cgatcggtcc ggtcggtggc tccccgcggc ggggccgggc 60
ccgatctcgg gcgggaaccg agcgcagagc cggtagcggg aaggatgacc acgctcacac 120
gacaagacct caactttggc caagtggtgg ccgatgtgct ctgcgagttc ctggaggtgg 180
ctgtgcatct catcctctac gtgcgcgagg tctaccccgt gggcatcttc cagaaacgca 240
agaagtacaa cgtgccggtc cagatgtcct gccacccgga gctgaatcag tatatccagg 300
acacgctgca ctgcgtcaag ccactcctgg agaagaatga tgtggagaaa gtggtggtgg 360
tgattttgga taaagagcac cgcccagtgg agaaattcgt ctttgagatc acccagcctc 420
cactgctgtc catcagctca gactcgctgt tgtctcatgt ggagcagctg ctccgggcct 480
tcatcctgaa gatcagcgtg tgcgatgccg tcctggacca caaccccca ggctgtacct 540
tcacagtcct ggtgcacacg agagaagccg ccactcgcaa catggagaag atccaggtca 600
tcaaggattt cccctggatc ctggcggatg agcaggatgt ccacatgcat gacccccggc 660
tgataccact aaaaaccatg acgtcggaca ttttaaagat gcagctttac gtggaagagc 720
gcgctcataa aggcagctga gggggcacct gccaccccac tgatgcccaa actgtcagac 780
tttgggggat ccccgcctag ggcagtgctg catggctgcc ctgattccaa gtgctcttat 840
cgcctctgtg tgtggatcgc ccgccccagc ccggggccgc tcaggtctgc ttggaggatg 900
cctcccccag gagggcagtg agggatgccg caacctcgac ttctcagcct cctggggttc 960
cgccggccaa cactgtctgt ctcaaatact gtgctgtgag ttgtttcaat aaaggggccc 1020
caagggctgg gctgaaaaaa aaaaaaaaaa aa 1052

<210> 7
<211> 2024
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Clone No: 2678913

<400> 7
agattgctct gatgtctttc cttccataag tcctacggag atcgagctta cctcattttt 60
ctcccaggcc taggagctca attaccagga tcactttctt ccccagggtg tactacagac 120
ctccacatgt tggacagtag gattcagcag agaatgagtc ccacatcctc ttttcatcca 180
taagccactc ttcccaacta ctcttccgcc ttccattgtc cctatgtcta ctttactgct 240
caatctggat tttggggaac ctcctcccaa aaaggcatta gaaggaaatg ccaagcaccg 300
aaattttgtc aagaagcgga ggctcttaga acggagaggc tttctgagta aaaagaacca 360
acccccctagc aaggcgccta agttgcactc tgaaccttca aagaaagggg aaactcctac 420
ggtcgatggc acttggaaga ccccttcctt cccaaaaaag aagacagctg cttccagcaa 480
tgggtcagga cagcccctgg acaagaaagc tgcagtgtct tggttgaccc ctgccccttc 540
aaaaaaggct gattctgttg ctgctaaagt agatttgctg ggggagttcc agagtgccct 600
tccaaagatc aatagccacc caacccgctc tcagaagaag agctcccaga agaaatcctc 660
taaaaagaac catcctcaga agaatgcccc acagaactcc acccaagctc attcagagaa 720
taaatgctcc ggagcatccc agaagttgcc acggaagatg gtggcaattg actgtgagat 780
ggtgggcaca ggaccaaagg ggcatgttag ttccttggct cgatgtagca ttgtcaacta 840
caacggagat gtgctttatg acgagtacat tcttcccccc tgccacattg tggactaccg 900

aaccaggtgg agtggtatcc ggaagcagca catggtgaat gccacaccct tcaagattgc 960
tcgaggccag atcttgaaga tactcacagg gaagatagtg gtggggcatg ccatccacaa 1020
cgacttcaaa gcccttcagt actttcaccc caagtccctc acccgtgaca cctcccatat 1080
ccccccccctc aaccggaagg ctgactgccc ggagaatgcc accatgtctc tgaagcatct 1140
caccaagaag ctgctaaacc gggatatcca ggttgggaag agcggacatt cctctgtgga 1200
agatgcccag gccaccatgg agctatataa gttggttgaa gtcgagtggg aagagcacct 1260
agcccggaat ccccctacag actagtggca gtggggacgc tggtgatatg aggaggcaga 1320
ggcagcaccc aggagaaaca gggcagtgga ccaatggaca gctccaccag ctccacatct 1380
ttggaagcta gatttgggga gagagaagct ctaccccaga cttaataccc attgaaattt 1440
cacctcaggt gttgtgtcct gtgtctggtt aagtgtccca tggaagggga aagccttcac 1500
gtcagaaccc aaccctatac cttttacttc ttaaatggtg ctaaccacag gtgtcccagg 1560
gtgctctgtg ccagttaaga tttttaactt tcaaggggca gggcatactg ggaaatgtag 1620
tttcccaaac tgccttatca cttgggtgga catatgtctc cttttatgcc ttttggtctt 1680
gagtaattaa cagcatcctc ttccacgctc agaagtgttc tggttggggc caggcatggt 1740
ggctcacgcc tgtagtccca acacttaggg aggccgaggc gggcggatca cctgagatca 1800
ggagttcaag accagcctgg ccaacatggc gaattcccgt tctctactaa aaatacaaaa 1860
aatgtgtggg gtgtggtggc aggagcctgt aatcctagct actcaggagg ctgaggcagg 1920
agaatcgctt gagcccagga ggcggagatt gcagtgagcc gagatcgtgt cactgcactc 1980
cagcctgggt gacaagagtg agactccgtc tccaaaaaaa aaaa 2024

<210> 8
<211> 2650
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Clone No: 1849911

<400> 8
taaaatcatt agtttggacc ttcctgtggc tgaagtttac aagaaagtct ggtgtaccac 60
gaatgaggga gagcccatga ggattgttta tcgtatgcgg gggctgctgg gcgatgccac 120
agaggagttc attgagtccc tggactctac tacagatgaa gaagaagatg aagaagaagt 180
gtataaaatg gctggtgtga tggcccagtg tgggggcctg gaatgcatgc ttaacagact 240
cgcagggatc agagatttca agcagggacg ccaccttcta acagtgctac tgaaattgtt 300
cagttactgc gtgaaggtga aagtcaaccg gcagcaactg gtcaaactgg aaatgaacac 360
cttgaacgtc atgctgggga ccctaaacct ggcccttgta gctgaacaag aaagcaagga 420
cagtgggggt gcagctgtgg ctgagcaggt gcttagcatc atggagatca ttctagatga 480
gtccaatgct gagcccctga gtgaggacaa gggcaacctc ctcctgacag gtgacaagga 540
tcaactggtg atgctcttgg accagatcaa cagcaccttt gttcgctcca accccagtgt 600
gctccagggc ctgcttcgca tcatcccgta cctttccttt ggagaggtgg agaaaatgca 660
gatcttggtg gagcgattca aaccatactg caactttgat aaatatgatg aagatcacag 720
tggtgatgat aaagtcttcc tggactgctt ctgtaaaata gctgctggca tcaagaacaa 780
cagcaatggg caccagctga aggatctgat tctccagaag gggatcaccc agaatgcact 840
tgactacatg aaaaagcaca tccctagcgc caagaatttg gatgccgaca tctggaaaaa 900
gtttttgtct cgcccagcct tgccatttat cctaaggctg cttcggggcc tggccatcca 960
gcaccctggc acccaggttc tgattggaac tgattccatc ccgaacctgc ataagctgga 1020
gcaggtgtcc agtgatgagg gcattgggac cttggcagag aacctgctgg aagccctgcg 1080
ggaacaccct gacgtaaaca agaagattga cgcagcccgc agggagaccc gggcagagaa 1140
gaaacgcatg gccatggcaa tgaggcagaa ggccctgggc accctgggca tgacgacaaa 1200
tgaaaagggc caggtcgtga ccaagacagc actcctgaag cagatggaag agctgatcga 1260
ggagcctggc ctcacgtgct gcatctgcag ggagggatac aagttccagc ccacaaaggt 1320
cctgggcatt tataccttca cgaagcgggt agccttggag gagttggaga ataagccccg 1380
gaaacagcag ggctacagca ccgtgtccca cttcaacatt gtgcactacg actgccatct 1440
ggctgccgtc aggttggctc gaggccggga agagtgggag agtgccgccc tgcagaatgc 1500
caacaccaag tgcaacgcgc tccttccggt ctggggacct catgtccctg aatcagcttt 1560

tgccacttgc ttggcaagac acaacactta cctccaggaa tgtacaggcc agcgggagcc 1620
cacgtatcag ctcaacatcc atgacatcaa actgctcttc ctgcgcttcg ccatggagca 1680
gtcgttcagc gcagacactg gcgggggcgg ccgggagagc aacatccacc tgatcccgta 1740
catcattcac actgtgcttt acgtcctgaa cacaacccga gcaacttccc gagaagagaa 1800
gaacctccaa ggctttctgg aacagcccaa ggagaagtgg gtggagagtg cctttgaagt 1860
ggacgggccc tactatttca cagtcttggc ccttcacatc ctgccccctg agcagtggag 1920
agccacacgt gtggaaatct tgcggaggct gttggtgacc tcgcaggctc gggcagtggc 1980
tccaggtgga gccaccaggc tgacagataa ggcagtgaag gactattccg cttaccgttc 2040
ttcccttctc ttttgggccc tcgtcgatct catttacaac atgtttaaga aggtgcctac 2100
cagtaacaca gagggaggct ggtcctgctc tctcgctgag tacatccgcc acaacgacat 2160
gcccatctac gaagctgccg acaaagccct gaaaaccttc caggaggagt tcatgccagt 2220
ggagaccttc tcagagttcc tcgatgtggc cggtctttta tcagaaatca ccgatccaga 2280
gagcttcctg aaggacctgt tgaactcagt cccctgacca ccacacagca gctgcggcgg 2340
cgaagacgaa gctggcttgc cttccaccct ctgttctccc tccttgtgca ttaagttccc 2400
tccgcgggat gctgcattgt taccccgccc tcccctctct catttttctt ggtgtggctt 2460
ggggttttta ggcttcctgt tttatctcgt gtgtgtggtg caccagctat gaggttgtct 2520
gtaacccaag ccatcaaagg gcctgtacat acctaggagc catgagttgt cccggccagc 2580
ttcatacttg agtgtgcaca tcttgagaaa taaacaagtg acttaacaca cattgaaaaa 2640
aaaaaaaaaa 2650

<210> 9
<211> 1921
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Clone No: 2223353

<400> 9
gcaggacggc gggaagagga gtgcggaacc cgcgggagag tccccaggga gacacttaag 60
ggaaattaaa ctgcagagtg caagagatgc ctcagtcaag tcagccaaaa acacgcgggt 120
catccccaag ccccagagag tgacagagcc ccgatgacac ggacacctcg gctgctgtca 180
cttccctggt tcgggcctcc cacaggcttt gaattgaagg cgagtgcctc agaatttgca 240
tccattgttc tgtctttcct gggaagttat tcatcctggt ggccagccca ccgacaaaat 300
ggatttggat ctactggacc tgaatcccag aattattgct gcaattaaga aagccaaact 360
gaaatcggta aaggaggttt tacacttttc tggaccagac ttgaagagac tgaccaacct 420
ctccagcccc gaggtctggc acttgctgag aacggcctcc ttacacttgc ggggaagcag 480
catccttaca gcactgcagc tgcaccagca gaaggagcgg ttccccacgc agcaccagcg 540
cctgagcctg ggctgcccgg tgctggacgc gctgctccgc ggtggcctgc ccctggacgg 600
catcactgag ctggccggac gcagctcggc agggaagacc cagctggcgc tgcagctctg 660
cctggctgtg cagttcccgc ggcagcacgg aggcctggag gctggagccg tctacatctg 720
cacggaagac gccttcccgc acaagcgcct gcagcagctc atggcccagc agccgcggct 780
gcgcactgac gttccaggag agctgcttca gaagctccga tttggcagcc agatcttcat 840
cgagcacgtg gccgatgtgg acaccttgtt ggagtgtgtg aataagaagg tccccgtact 900
gctgtctcgg ggcatggctc gcctggtggt catcgactcg gtggcagccc cattccgctg 960
tgaatttgac agccaggcct ccgcccccag ggccaggcat ctgcagtccc tgggggccac 1020
gctgcgtgag ctgagcagtg ccttccagag ccctgtgctg tgcatcaacc aggtgacaga 1080
ggccatggag gagcagggcg cagcacacgg gccgctgggg ttctgggacg aacgtgtttc 1140
cccagccctt ggcataacct gggctaacca gctcctggtg agactgctgg ctgaccggct 1200
ccgcgaggaa gaggctgccc tcggctgccc agcccggacc ctgcgggtgc tctctgcccc 1260
ccacctgccc ccctcctcct gttcctacac gatcagtgcc gaaggggtgc gagggacacc 1320
tgggacccag tcccactgac acggtggcgg ctgcacaaca gccctgcctg agaagccccg 1380
acacacgggg ctcgggcctt taaaacgcgt ctgcctgggc cgtggcacag ctgggagcct 1440
ggttcagaca cagctcttcc agggcagcgg ctccactttc tcatccgaag atggtggcca 1500
cagactgacc cccatctgag ctgggggggat gttctgcctc tccctgggtc tggggacagg 1560

cccgcttgct gggtacctgg tccccactgc tgagctggcc cttggggaga ggtgattctc 1620
agggctggag cctggggtgt cctacagtga ctccctggga gccgcctgct tcttctctcc 1680
acatggaagc ccaactgggg ttgcgtctga ggcctgcccc ctgggctggg gcctcagacc 1740
ccctcagcct tgggaccgtg cccacgaggg tctcccctcc tgcacacagg gcagtcctta 1800
ctcccccacc actcaggcca cagtggggct gcaggcaggc ggctcctcct cacccacctc 1860
tgggtccttg gctcccgggg gccccacttc ggcacacact gtgccccaca aaattcagtg 1920
a 1921

<210> 10
<211> 1283
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Clone No: 2593035

<400> 10
tgggtaaccg gggatcacca tggcggcctc attggtgggg aagaagatcg tgtttgtaac 60
ggggaacgcc aagaagctgg aggaggtcgt tcagattcta ggagataagt ttccatgcac 120
tttggtggca cagaaaattg acctgccgga gtaccaaggg gagccggatg agatttccat 180
acagaaatgt caggaggcag ttcgccaggt acaggggccc gtgctggttg aggacacttg 240
tctgtgcttc aatgcccttg gagggctccc cggcccctac ataaagtggt ttctggagaa 300
gttaaagcct gaaggtctcc accagctcct ggccgggttc gaggacaagt cagcctatgc 360
gctctgcacg tttgcactca gcaccgggga cccaagccag cccgtgcgcc tgttcagggg 420
ccggacctcg ggccggatcg tggcacccag aggctgccag gactttggct gggacccctg 480
ctttcagcct gatggatatg agcagacgta cgcagagatg cctaaggcgg agaagaacgc 540
tgtctcccat cgcttccggg ccctgctgga gctgcaggag tactttggca gtttggcagc 600
ttgacttctg cagctggagg aggcccctca ggccggggat ctggggaggg ctagcccaaa 660
acctcccgca tcgggcaggc acccccctgaa gtacttcctt cagggtttcc cctttgtgag 720
ggtgtcgagt agcctcaccg gcctgtctgg aggagcagct ggctctgctc tgagaaactc 780
tggcaagtgg acgccattct cttgccctta ggattcactg ctctctccta cagccgccag 840
gcctggggtc ctgaaaggac cttgggtggt aaagctgtac ttggtgggag tgagggcgtg 900
gggaggaacc atgcaaatcg ccttccatgg tttttaaatg cagtaaataa catttctgga 960
tgagacttgt ttccaaaata aaccagctat atctgttttg aaaaacaaat ccaccaagtc 1020
ctaaaaaaag aatcacaatg acggcaagaa attcaaagag cccttaaagg caatcccccct 1080
agggaaacct cttgtaacca aaactggggg cccccgcata aggcacccca tgcccggggt 1140
ttttattccg ggacgggttc ccttaggggg tcccgggttt ccccctagag gggggggttt 1200
taaaacttct gggggaactc aggcccaaaa gtgttctccc gggggggaaat tcgtttcccg 1260
cccccaaatt tctcccccccc acc 1283